

Weekly Report 2017.09.11-2017.09.17

Progress

1. 大图可视化调研

2. 可视化专著

3. RCanalyzer 修改

首先，本周与周大为进行了讨论，跟他确定了需要他来撰写的部分，包括算法的描述和算法的评估两个部分。大为说他已经做好了一个初始的版本，正在进行测试，但是需要具有属性的动态图。我们之前生成的数据是包含属性的，因此我打包了一份数据发给了他，目前他应该在进行测试中。

其次，我已经完成了第三章的撰写（对 **rare category** 的描述），第四章算法部分交给大为去写。我目前已经开始撰写论文的第五章，即可可视化部分。

最后，因为嘉铨和东明都去深圳了，所以系统开发的工作我接过来做了一部分，主要是新数据的导入和读取。我重新抽取了一份大约有 6000 个节点的数据，并导入到了 neo4j 中。之前遇到了 neo4j 无法分表的问题，即两个不同的网络存在 neo4j 中是混在一起存的。我目前采用的解决方案是给所有的节点加一个 **dataset** 字段，查询时加上这个字段来对网络进行区分。同时，由于 6000 个节点的图的邻接矩阵就有 144m 大，在动态图的情况下内存扛不住，因此我将底层的数据存储转到了 scipy 的 **csrmatrix** 结构中，这是一种专门针对稀疏矩阵设计的数据结构，同时支持快速的矩阵运算。目前新的数据已经可以正常运行加载了，但是由于新的 **bird** 算法还没有做好，因此还没有办法看到效果。

工作	截止日期	当前进度
RCanalyzer 文章修改	-	最重要的一章已经写完，开始撰写可视化这一章。新数据的存储和加载问题已经解决，后面更大数据量应该也可以 hold 得住